MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 1014–1027 doi: 10.1111/mec.12656

Local and regional founder effects in lake zooplankton persist after thousands of years despite high dispersal potential

M. VENTURA,*† A. PETRUSEK,‡ A. MIRÓ,* E. HAMROVÁ,‡ D. BUÑAY,* L. DE MEESTER§ and J. MERGEAY§¶

*Center for Advanced Studies of Blanes, Spanish Research Council (CEAB-CSIC), Biodiversity and Biogeodynamics Group, Blanes 17300, Catalonia, Girona, Spain, †Institut de Recerca de l'Aigua, Universitat de Barcelona, Av. Diagonal 684, 08034 Barcelona, Catalonia, Spain, ‡Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, CZ-12844 Prague 2, Czech Republic, §Laboratory of Aquatic Ecology, Evolution and Conservation, KU Leuven, University of Leuven, Deberiotstraat 32, B-3000 Leuven, Belgium, ¶Research Institute for Nature and Forest, Gaverstraat 4, B-9500 Geraardsbergen, Belgium

Abstract

We reconstructed the genetic structure of a planktonic crustacean *Daphnia longispina* living in high mountain lakes and ponds in the Pyrenees to investigate whether it was shaped by persistent founder effects originating shortly after the last glacial maximum or by ongoing dispersal and effective migration (gene flow). We found that the genetic structure can largely be explained by a single colonization event following gradual deglaciation of the Pyrenees ~10 000–15 000 years ago. Nuclear genetic diversity declined steeply from southeast to northwest, suggestive of serial colonization of available habitats with advancing deglaciation. The spatial genetic structure suggests that founder effects were major determinants of the present-day diversity, both at the catchment level and at the level of individual water bodies, further supporting extremely low effective migration rates. This study reveals a prime example of a founder effect that is both long lasting and maintained at small spatial scales. Our data suggest a process of isolation by colonization as a result of strong priority effects and monopolization. We found evidence for the spread of haplotypes with Pyrenean ancestry across the Palaearctic over distances up to 5500 km, although the local genetic structure after colonization was hardly influenced by contemporary dispersal. Finally, our data also suggest that mitochondrial mutation rates in the studied populations were seven times higher than typically assumed. Overall, we show that founder effects can persist for centuries even at small spatial scales at which the potential for dispersal is high.

Keywords: alpine lakes, *Daphnia*, dispersal–geneflow paradox, founder effects, monopolization hypothesis, Pyrenees

Received 20 September 2013; revision received 22 December 2013; accepted 30 December 2013

Introduction

The relative importance of neutral vs. selective forces in determining local genetic differentiation in natural populations is a major question in contemporary evolutionary ecology, which is subject to much debate (Hey 1999). Understanding species distributions and the relevance of environmental factors affecting them, relative to spatial variation, requires an understanding how local environmental factors affect the dynamics of local populations and structure of metapopulations in relation to regional spatial constraints. Local adaptation can result from immigration of preadapted individuals and further spread of the traits in the rest of the population, that is, lineage sorting (Wade 2000), or from standing

Correspondence: Marc Ventura, Fax: +34972337806;
E-mail: ventura@ceab.csic.es

genetic variation present in the founding population (Barrett & Schluter 2008; Orsini *et al.* 2013a). Whereas the first mechanism is expected to cause a pattern of genetic isolation by environment for neutral markers [or environmental genetic structuring, see Nosil *et al.* (2008)], the latter is expected to reflect colonization patterns, which may result in a pattern of isolation by distance (cf. concept of isolation by colonization proposed by Orsini *et al.* 2013a). These two mechanisms provide additional explanations to population genetic differentiation at the landscape level to the traditional neutralist view of the isolation-by-dispersal limitation, which typically leads to a pattern of isolation by distance (see Orsini *et al.* 2013a). Which of these mechanisms is most likely to act in natural populations depends on the rate of effective gene flow into the population, the amount of standing genetic variation for the traits under selection and the increase in genetic variation through mutations.

In many species, assumptions about dispersal rates based on ecological insights strongly contrast with geneflow estimates. A good example is freshwater zooplankton, for which studies on the rapid colonization of new habitats and spread of invasive species suggest a high potential for dispersal (Havel *et al.* 2000; Louette & De Meester 2004; Mergeay *et al.* 2006), but genetic studies show that the realized gene flow is often much lower (Boileau *et al.* 1992; De Meester 1996). This dispersal–geneflow paradox can be explained by founder effects (Boileau *et al.* 1992), further enhanced by local adaptation, thereby causing monopolization of resources by local populations (De Meester *et al.* 2002). This process leads to a pattern in which population genetic variation at the landscape level is correlated with the pattern of colonization (isolation by colonization; Orsini *et al.* 2013a). Insight into the importance of realized gene flow among populations is central to answering whether adaptation occurs through gene flow and lineage sorting, or through standing genetic variation and local adaptation. This information can partly be acquired by studying the distribution of neutral genetic variation among populations on different temporal and spatial scales, for example, by combining phylogeography with population genetics at hierarchical spatial scales.

Here, we focus on the genetic diversity and distribution of *Daphnia longispina* (O.F. Müller, 1776), a widespread and ecologically plastic Palaearctic species of freshwater zooplankton (see Petrusek *et al.* 2008), which we studied across lakes and ponds of the eastern Pyrenees. *Daphnia* are keystone species in freshwater habitats, being grazers of suspended algae and prey for vertebrate and invertebrate zooplanktivores. Although this is certainly not the first study on the genetic diversity and distribution of *Daphnia* in high mountain ranges (e.g. Aguilera *et al.* 2007; Petrusek *et al.* 2007; Mergeay *et al.* 2008; Hamrová *et al.* 2012), it is noteworthy that comparatively little attention has been paid to the ecology and evolution of high mountain zooplankton compared to lowland populations (Wetzel 2001). This is despite the fact that high mountain regions can be considered important reservoirs of genetic variation, given the often complex response of biota to past climate changes, steep ecological gradients found over small to large spatial scales and resulting complex patterns of glacial refugia (Hewitt 2000).

Like most high mountain ranges, the Pyrenees have a varied spatial structure (Catalan *et al.* 2009). These mountains are dotted with ~4000 lakes and ponds of glacial origin, dating from the late Pleistocene (Catalan *et al.* 2006). Deglaciation during the last glacial period occurred markedly earlier in the Pyrenees (15 000–10 000 years ago, in the altitudes where glacial lakes are situated) than in the Alps and northern Europe (Delmas *et al.* 2008), making Pyrenean glacial lakes slightly older than others in Europe. The whole mountain range of the Pyrenees has been defined, based on lake species, as a single ecoregion (Catalan *et al.* 2009), which makes it a suitable area for studying species distribution at the metapopulation level. In addition, the island-like nature of limnetic habitats, and the particular nonhomogeneous distribution of Pyrenean lakes with various groups of lakes distributed along the mountain range, is expected to create higher opportunities for local genetic differentiation and adaptation to develop. Moreover, *D. longispina* occurs throughout the Pyrenees both in very small shallow ponds and in larger deep lakes across different catchments. Such habitats differ enormously (e.g. ponds freeze entirely during winter and many dry over the summer), and likely also require very different adaptations, to the point that many speciation events in *Daphnia* have been attributed to shifts from lakes to ponds and vice versa (Lynch 1985; Lynch & Spitze 1994). This study provides an opportunity to investigate whether or not shifts from lakes to ponds in the Pyrenees have occurred only once (with subsequent environmental lineage sorting across the region) or whether it has occurred independently at multiple occasions across catchments.

Our general aim was to study the relative importance of environmental vs. spatial structure in determining local genetic structure in Pyrenean high mountain lake populations of *D. longispina*, and to assess which is the most likely mechanism causing population genetic differentiation: isolation-by-dispersal limitation, isolation by adaptation or isolation by colonization. Specifically, we determined genetic diversity and differentiation at a mitochondrial marker (12S rDNA) and at nine nuclear

microsatellite loci. Whereas microsatellites provide information on relatively recent demographic processes, mitochondrial DNA is more suited to reconstruct more ancient patterns of colonization and dispersal (phylogeography). From the spatial distribution of haplotypes among waters within and among catchments, we can make inferences about the importance of dispersal as an ecological process at three spatial scales: the whole sampled region (eastern Pyrenees), at the catchment level and at the individual lake level. We then combined this genetic information with lake characteristics to analyse whether adaptation to local selection factors occurred and whether they were most likely driven by gene flow (lineage sorting) or by parallel local adaptation. In this way, we can differentiate among isolation-by-dispersal limitation, isolation by adaptation or isolation by colonization (Orsini *et al.* 2013a).

Materials and methods

Study area and sample collection

Zooplankton samples were collected from 25 alpine lakes and ponds in the eastern (Catalan) Pyrenees (Fig. 1 and Table S1, Supporting Information) during the lake ice-free periods from 2005 to 2008. Samples were obtained by vertical net hauls from the surface with a 200-μm net and preserved in absolute ethanol for genetic analysis. We sampled nine different catchments and attempted to sample multiple lakes (2–6) per catchment. In two cases, the catchment consisted of a single lake. Often, water bodies within a catchment were interconnected or were located close to each other (see Table S1, Supporting Information for further details). Such connectivity levels allowed us to calculate expected passive dispersal rates of adults across interconnected lakes.

For each sampled lake, the following environmental variables were recorded: area, maximum depth, presence of potential predators (fish, *Gammarus lacustris*, *Cyclops abyssorum*), total nitrogen content (TN), total phosphorus content (TP), pH, alkalinity, conductivity, concentrations of major cations (K, N, Mg, Ca), dissolved organic carbon (DOC) and dissolved inorganic carbon (DIC). The sampling procedures and analytical methods are explained in detail elsewhere (Ventura *et al.* 2000).

Genetic analysis

DNA was extracted from 30 to 60 *Daphnia longispina* individuals per population by proteinase K digestion (Schwenk *et al.* 1998) in 100 μL volumes. One microlitre of DNA extract from these samples was used in subsequent polymerase chain reactions (PCR).

Per population, DNA from ~10 individuals was subjected to DNA sequencing to provide a reasonable coverage of haplotype diversity within and among lakes. We amplified ~540 nucleotides (nt), a long fragment of the mitochondrial 12S rRNA gene (12S) following standard protocols (e.g. Petrusek *et al.* 2008; Appendix S1, Supporting Information). Purified PCR products were sequenced using forward primers on an ABI PRISM 3130 capillary DNA sequencer (Applied Biosystems).

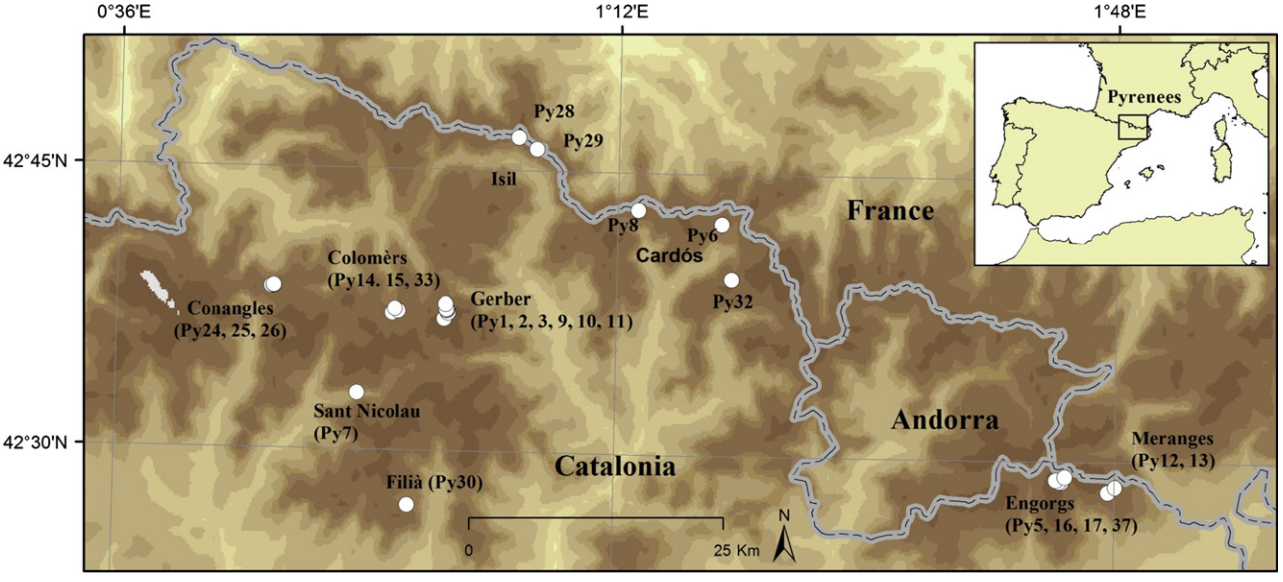


Fig. 1 Distribution of sampling sites along the eastern (Catalan) Pyrenees. White circles represent each of the 25 water bodies sampled; names indicate the different catchments. See Table 1 for the basic lake characteristics.

We failed to amplify sequences from six water bodies due to low template DNA quality and the sequences could not be obtained due to unavailability of other samples at the time of analysis. The resulting 127 sequences of 528 nt length were aligned with 417 sequences of other available *D. longispina* populations from the known species distribution range (Table S2, Supporting Information) using the CLUSTALW algorithm in MEGA 5 (Tamura *et al.* 2011).

Nine microsatellite loci were amplified in a single multiplex PCR in 10 μL volume consisting of 5 μL HotStar *Taq* DNA polymerase buffer (Qiagen, Hilder, Germany), 0.4 μM of each primer for locus SwiD15, 0.3 μM for loci SwiD1, SwiD2, SwiD12, SwiD14, Dpu6 and Dgm109, 0.2 μM for loci SwiD10 and Dp196NB and 0.05 μM for Dp281NB (Brede *et al.* 2006). Cycling conditions were 15 min at 95 °C followed by 30 cycles of 30 s at 94 °C, 90 s at 54 °C and 60 s at 72 °C and a final elongation step at 60 °C for 30 min. Polymorphism was assessed on an ABI PRISM 3130 capillary DNA sequencer, using an internal Liz Gene-scan size standard (Applied Biosystems). All primer pairs amplified unambiguous PCR products between 69 and 266 nucleotides long. Tests of data quality were performed using MICRO-CHECKER (Van Oosterhout *et al.* 2004), and all markers passed the quality test with absence of null alleles at all loci.

Genetic variation of mitochondrial and nuclear markers and demography inferences

Haplotype and nucleotide diversity levels per lake for mtDNA were calculated using DNASP 5.0 (Rozas *et al.* 2003). We performed a network analysis to estimate gene genealogies using HAPLOVIEWER, which turns trees built from traditional phylogenetic methods into haplotype genealogies (Salzburger *et al.* 2011). We estimated the phylogeny using a maximum-likelihood method with RaxML Blackbox (Stamatakis 2006), with gamma model of rate heterogeneity (GTR + G) and no invariant sites as suggested by the model. Input data were 12S sequences from each individual, subsequently collapsed into haplotypes. Sequences with ambiguous bases were not included in the analysis. The best tree (using the log-likelihood criterion) was selected for network construction using HAPLOVIEWER.

For microsatellite loci, population genetic parameters such as deviations from Hardy–Weinberg and linkage equilibria, inbreeding coefficients and the number of private alleles were calculated using programs GENEPOP 4.2 (Raymond & Rousset 1995) and GENALEX 6.5 (Peakall & Smouse 2006). Estimation of clonal diversity and genetic differentiation was based on several population genetic parameters, such as the number and relative abundances of distinct multilocus genotypes (MLG), Simpson's diversity index and a standardized measure of allelic richness to a constant sample size (Szpiech *et al.* 2008). We took the minimum number of individuals analysed in a population as the sample size for all populations (29), to standardize the richness measure.

For mtDNA, we tested the hypothesis of a population expansion event by calculating Tajima's *D* (Tajima 1989) with 10 000 permutations in ARLEQUIN 3.5 (Excoffier & Lischer 2010). Negative values in neutrally evolving genes suggest deviations from mutation–drift equilibrium due to population expansion events. Next, we calculated the mismatch distribution among Pyrenean haplotypes (Rogers & Harpending 1992), under the spatial expansion model. The sudden expansion model was not considered for our data set because it is less likely to represent the real situation than a spatial expansion model, given the spatially structured nature of lake populations and the gradual deglaciation of the Pyrenees from SE to NW. The mismatch distribution characterizes the expected number of pairwise differences among sequences given a certain model of population change. ARLEQUIN 3.5 thus simulates, given the data, the expected pattern under a spatial expansion model and estimates whether or not the actual data significantly differ from the simulated distributions. This analysis gives the demographic parameter tau, which was used to estimate the timing of the clade expansion (t), using the equation $\tau = 2m_T\mu t$, with m_T being the number of nucleotides (528) and μ the mutation rate of the 12S gene per generation. For high mountain ultraoligotrophic lake *Daphnia pulicaria*, one generation per year has been described (Ventura & Catalan 2005), although in north temperate lakes other authors have inferred five generations per year (Costanzo & Taylor 2010). In order to confirm the number of generations per year in our study species, we sampled one of the studied lakes (Lake Llarg; Fig. S2, Supporting Information) with a relatively higher nutrient content (thus being at the upper range within our study sites) during two consecutive ice-free seasons (2010–2011). The lake was sampled at monthly intervals at the deepest point of the lake by vertical hauls, and lake *Daphnia* were counted and measured following standard procedures (Ventura & Catalan 2005) in order to obtain an estimate of the number of generations.

The experimental quantification of mtDNA mutation rate of *Daphnia* has recently been quantified to be among the highest in eukaryotes [ranging from 14% to 17% per million (M) generations; Xu *et al.* 2012]. This mutation rate is similar to those of other invertebrates such as *Caenorhabditis elegans* (16% M generations; Denver *et al.* 2000), but substantially higher than those estimated for the mitochondrial gene cytochrome *c* oxidase subunit I (COI), and used in many studies (2.3%

M generations; Brower 1994) or those used for north temperate *Daphnia* (6.6% M generations; Costanzo & Taylor 2010). 12S could be assumed to be at the lower end of the mutation rates found in mtDNA. This is justified by the comparison with COI, for which a direct comparison of its variation with the 12S rRNA gene in *Daphnia* reveals that at shallow levels of genetic divergence, the 12S gene is ~1.7 times less variable than COI (J. Mergeay, unpublished). In *Daphnia* living in environments with high UV stress, mutation rates have been found to be ~2.9 times higher than normal (Hebert *et al.* 2002). All *Daphnia* in the study lakes and ponds are melanized, supporting that assumption that the UV stress is high at these high-altitude habitats (Laurion *et al.* 2000). Since the date of melting of the upper glaciers of the Pyrenees (where the high mountain lakes are located) is known (ranged between 15 000 and 10 000 years ago; Delmas *et al.* 2008), we used this advantage of our study system to test which mutation rate was needed to explain the measured diversification rates, assuming that mutations occurred *in situ* (see below). Therefore, we simulated different colonization times based on the timing of clade expansion for a range of mutation rates from 1.5% to 24% per million generations and different number of generations (from 1 to 5).

We estimated the number of independent colonization or dispersal events at three spatial scales (based on the sharing of observed haplotypes or the most parsimonious ancestral ones deduced from the haplotype network). This was done at the regional level (Pyrenees vs. rest of the species range), at the catchment level within the Pyrenees and at the water body level within a catchment by performing analysis of molecular variance (AMOVA) at these three levels using ARLEQUIN 3.5, separately for mtDNA and nDNA data. For mtDNA, we used the Kimura 2-parameter distance among haplotypes to infer genetic differentiation (F_{ST}).

Disentangling the drivers of population genetic differentiation

To examine genetic patterns caused by spatial structure that would reflect of isolation by distance, we performed a redundancy analysis (RDA) using distance-based Moran eigenvector maps (MEM) as independent variables and the nuclear genetic data as dependent variables (Borcard *et al.* 2004; Dray *et al.* 2006). This approach is much more powerful to detect spatial structure and the scale of it than more commonly used tests such as Mantel tests between pairwise geographical and genetic distances (Legendre & Fortin 2010). We calculated Nei's genetic distance determined at nine microsatellite loci in GENALEX between all populations, and this distance matrix was used to calculate principal coordinates (PCo) of the genetic data in the R package Vegan (Oksanen *et al.* 2010). This approach is preferable to using the allelic frequencies directly as dependent variables, because of the intrinsic dependency of alleles within a locus, but not among loci. The MEM were constructed using the spatial coordinates of each sampled lake or pond. In addition, three evenly spaced dummy points were added between the lake PY32 and the pond PY5 because of a relatively large unsampled area between them. Adding a small number of dummy samples can help to construct better spatial models when sampling designs are irregular (Borcard & Legendre 2002). The MEM were calculated by determining the minimum-spanning Euclidean distances among all sampling sites and by truncating distances larger than the largest minimum-spanning distance to four times this distance (Borcard *et al.* 2011). This truncated matrix was then subjected to a PCo analysis, which yielded five positive axes (MEM variables). They are ordered to represent a gradient from large-scale spatial structure (approximately among catchments and between regions) to small-scale structure (approximately within catchments). From such data, the relative importance of small and large spatial scales in the genetic data can be determined during a forward selection procedure in RDA (Borcard *et al.* 2004), by estimating which variables contribute most to the overall explained variance. Adjusted R^2 values were calculated to account for random correlations when the number of explanatory variables is large compared to the number of dependent variables. During the forward selection procedure, we used Blanchet's procedure to avoid the problem of inflation of the overall type I error and to reduce the risk to incorporate too many variables in the model (Blanchet *et al.* 2008).

In addition, we determined the contribution of environmental structure to the genetic variation in a similar way to verify to what extent there was isolation by environment. We used a subset of the measured environmental variables, selected by a principal components analysis of all measured environmental variables. Whenever correlated variables clearly overlapped in the analysis (e.g. Mg-Ca conductivity, lake area and lake depth), a single variable was retained. Hence, 12 variables were kept in the full environmental model: the presence of three predators (*Gammarus*, *Cyclops* and fish), lake size (log-transformed), pH, TP, TN, DIC, DOC, Na, K and conductivity. To disentangle the contributions of space and environment, we partitioned the genetic variance into purely environmental (*E*) and spatial (*S*) components, following the procedure explained in Peres-Neto and Legendre (2010). We first tested whether the overall model of each matrix (*S* or *E*) was significant. If so, we used forward selection of variables to produce a more parsimonious model. Only

significant variables ($P < 0.05$) were retained in further analyses and for variance partitioning. RDA and variance partitioning were performed according to the R package Vegan. We estimated the proportions of spatially structured genetic variation (S), environmentally structured variation (E), overall structure ($S + E$), exclusively spatial ($S \mid E$) and environmental ($E \mid S$) components and their overlap (shared spatial and environmental variance).

Isolation by distance is typically interpreted as the result of additive effects of continuous gene flow, but can also result from serial founder effects (Orsini *et al.* 2013a). Such priority effects along a linear path of colonization have been observed in a number of organisms (Ramachandran *et al.* 2005; e.g. Clegg *et al.* 2002; Duvernell *et al.* 2008), but typically on large spatial scales. In the Pyrenees, we hypothesized that a decrease in genetic diversity following the retreat of the Late Pleistocene glaciers (~10 000–15 000 years ago) from ESE to WNW (a surrogate of the age gradient of the lakes) would be in agreement with expectations under a scenario of serial founder effects and thus reflect a case of isolation by colonization. To test for such relationship, we performed a linear multiple regression analysis between nuclear gene diversity (H_e) as dependent variable and longitude and latitude as independent variables in the R package Vegan. We performed a regression analysis with sequential forward variable selection. Lake size is expected to affect gene diversity, as it is a proxy for population size, which in turn affects the equilibrium gene diversity. We therefore included lake size (log-transformed) as a covariable when regressing the geographical variables against H_e, to estimate the contribution of geography to genetic diversity irrespective of lake size.

Results

Genetic diversity

The alignment of 127 Pyrenean sequences of mtDNA contained 64 variable nucleotides of a total of 531. In total, we found 61 haplotypes in the Pyrenean lakes and ponds, all belonging to *Daphnia longispina*. No other *Daphnia* species or cryptic lineages were found in the sampled lakes. The remaining 417 sequences, mostly European but including also samples from Siberia, the Middle East and Africa, had 97 variable nucleotides, with 118 haplotypes. The average nucleotide diversity from the Pyrenees (Pi) was 0.0099, very similar to those from the Old World (Pi = 0.01062; Appendix S1, Supporting Information).

With regard to the variation in nuclear DNA, we recorded a total of 132 alleles at nine microsatellite loci, yielding an average of 14.7 alleles per locus overall (min: 5 at Dp196NB; max: 24 at Dgm109). At the population level, the average number of alleles per locus was much lower, from 2.3 in Py26 to 6.9 in Py8 (see Table S3 and Appendix S1, Supporting Information for detailed results).

Timing of colonization and expansion

The network analysis of the 12S mtDNA (Fig. 2a) shows that no single haplotype was shared between the Pyrenees and populations analysed elsewhere, although several European or African haplotypes differed only by a single point mutation from Pyrenean haplotypes. Moreover, the haplotype variation indicates that almost all sequences from the Pyrenees likely had a single local ancestor (Fig. 2a). The exception were two related sequences from Py12 for which most closely related haplotypes were found in the Carpathians (Tatra Mountains, ~1500 km away) and Russia (2900–5500 km). In reverse, we found non-Pyrenean haplotypes that have the most parsimonious ancestor from within the Pyrenees. These include temporary ponds in southern Spain (Zahillo, Sopetón and Taraje; all in Doñana National Park at ~800 km), lakes and ponds in Switzerland (Arosa, 760 km and St-Bernard, ~560 km), a reservoir in Czechia (Vranov, ~1300 km) and lakes in Slovakia (Dankovo, ~1500 km) and Russia (Glubokoje, ~2900 km and Chany Lake basin, ~5500 km).

Tajima's D yielded a significantly negative value of −1.81 (Table 1), suggesting a deviation from mutation–drift equilibrium indicative of a historical population expansion event. The mismatch distribution was unimodal, evidencing an expansion from a single ancestral source (Fig. S1, Supporting Information), showing an excellent fit of the observed data to the spatial expansion model. Parameters of the mismatch distribution and the estimated time of divergence are given for all samples, and for the eastern and western halves of the sampling area, in Table 1.

The seasonal study in Py17 showed that most of the individuals of *D. longispina* spend the ice-cover period as resting stages, whereas the number of generations per year did not exceed three (one sexual generation associated with the production of the ephippia, and one or two asexual generations) in the two studied consecutive years (Fig. S2, Supporting Information). The simulation of the different colonization times as a function of mutation rate at various generations per year revealed that only at mutation rates higher than 8% per million generations the results fit the known postglacial ice-melting period (Fig. 3) and a scenario of a single founding event for most Pyrenean haplotypes (as indicated by the haplotype network). Assuming that three

generations per year are typical for the studied lakes, a mutation rate of 14% per million generations (the minimum mutation rate measured by Xu *et al.* 2012) yields an estimated time since divergence (and hence of colonization of the eastern Pyrenees) of 8000 years ago (4000–15 000 years ago), and the difference of colonization time between the eastern and western areas was 3400 years ago (Table 1).

Dispersal within the Pyrenees inferred from mtDNA

The haplotype network revealed a high degree of endemicity of Pyrenean haplotypes and haplotype groups, and very remarkable patterns of genetic isolation despite direct connectivity of some habitats (Fig. S3, Supporting Information). Only six haplotypes were shared between two or more sampled water

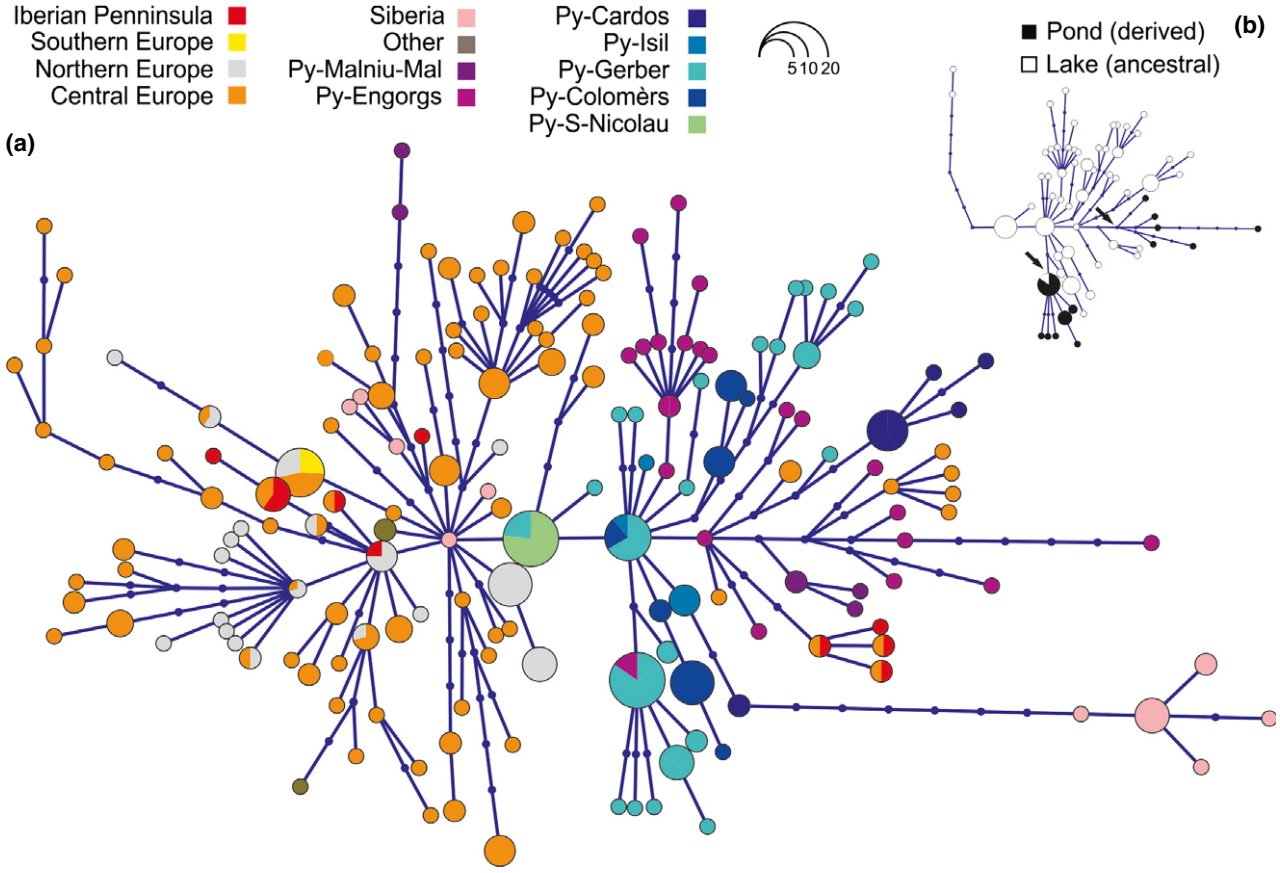


Fig. 2 Network of haplotypes of the mitochondrial gene for 12S rRNA. Each circle represents a unique haplotype, and its size is proportional to the number of individuals sharing that specific haplotype. Each branch with more than one mutational step is labelled. (a) Pyrenean catchments are indicated by different colours, as well as haplotypes from the rest of the species' range. The small insert (b) is a simplified network showing whether the haplotype was found in a lake (white circle) or a pond (black circle) or both (mixed); related haplotypes from the same water body or catchment were collapsed into an ancestral haplotype. Shifts from lake to pond or vice versa have occurred at least two times in the Pyrenees (black arrows).

Table 1 Estimated parameters of the mismatch distribution (spatial model for the Pyrenees, demographic expansion model for the individual clades). The mutation rate is set at 11% per million generations, and the number of generations per year is set at 2. CI, confidence interval at the 95% level

Clade	Tajima's *D*	*P*	Fu	*P*	Tau est.	*P*	Tau 95% CI	Age (years ago)	Age CI
Pyrenees	−1.81	0.008	−25.40	<0.001	3.56	0.952	1.85–6.48	8027	4171–14 610
Engorgs–Meranges lakes	−1.53	0.05	−21.41	<0.001	4.24	0.117	2.87–6.49	9560	6471–14 543
Colomèrs–Gerber lakes	−1.45	0.05	−16.86	<0.001	2.71	0.878	1.48–5.20	6110	3089–11 905

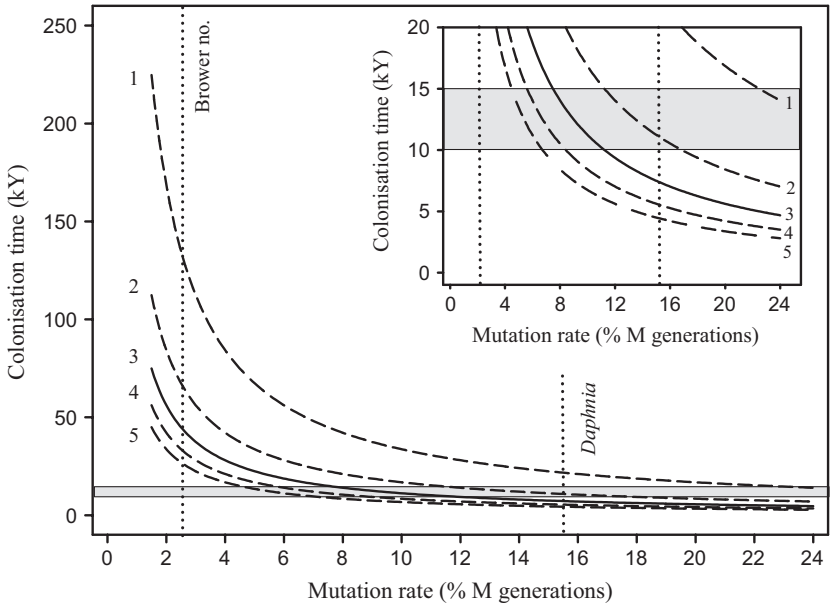


Fig. 3 Modelling of *Daphnia longispina* colonization time as a function of different mutation rates and number of generations per year given the observed diversity in haplotypes. Each line represents the colonization time calculated for 1–5 generations per year (numbers in the graph); three generations per year are highlighted by a solid line. Grey area marks the period when upper glaciers melted within this area of the Pyrenees (Delmas *et al*. 2008); lines above this area are thus erroneous estimates. Vertical dotted lines indicate the Brower number (Brower 1994) and the experimentally measured *Daphnia* mutation rate (Xu *et al*. 2012). The small panel is a closer view to the period when the ice melt started in the Pyrenees.

bodies. In three cases, these were shared between interconnected waters, whereas on the remaining eight occasions these were shared among nonconnected waters (three within catchments, five among different catchments; see Appendix S1, Supporting Information for more details). We found only sparse evidence for multiple colonization/dispersal events per catchment or per lake. The average number of independent colonization/dispersal events at the catchment level was estimated at 1.71 (Gerber: 2; Colomèrs: 2; Cardós: 2; Engorgs: 2; Meranges: 2; Isil: 1; Sant Nicolau: 1). At the pond or lake level, this trend was further confirmed, with an average of only 1.1 independent ancestral haplotypes per water body (Table S1, Supporting Information).

Spatial vs. environmental structure

The patterns of nuclear genetic variation revealed by PCo (Fig. 4B) showed a clear separation into two main geographical groups: the most eastern lakes (Meranges, Engorgs and Cardós) from the rest. Moreover, populations were generally clustered according to their catchment. Ponds and lakes from the same catchment were grouped together, not showing genetic segregation according to habitat type, indicative of independent habitat shifts through local adaptation instead of lineage sorting.

The analysis of molecular variance (AMOVA) revealed a strong genetic structure among catchments, with nearly 30% of the total genetic variation explained by catchment for the microsatellite loci and 21% for the mitochondrial DNA, compared to 10% and 46%, respectively, among populations within catchments; all values were highly significant ($P < 0.0001$; Table 2). When this analysis was repeated with populations grouped according to habitat type (lake vs. pond), only 2.15% ($P = 0.134$) of the diversity was explained by this grouping, compared to 35.6% explained by among-population variation within lake/pond groups.

In the redundancy analyses (Table 3), four of five spatial MEM variables were selected in the spatial model (S), explaining 67.1% (adjusted R^2) of the genetic variation, and only the intermediate MEM3 was not retained. After taking covariation with environmental variables into account, purely spatial variation ($S \mid E$) still explained 47.2% of the genetic variance. Similar to the AMOVA results, small-scale spatial variation (MEM4-5) explained a smaller proportion of the genetic structure than large-scale spatial variation (MEM1-2). In the environmental RDA matrix, only *Gammarus* and lake area were retained after forward selection, explaining 20.1% of the genetic structure. When space was included as a covariable matrix ($E \mid S$), no residual variation was explained. Combined ($S + E$), the forward selected spatial and environmental variables explained 70.1% of the genetic structure. The spatially structured environmental variance (shared) amounted to 22.9%, largely reflecting the fact that *Gammarus* only occurs in certain eastern lakes.

In a sequential multiple regression analysis, longitude conditionally explained 60.7% (adjusted R^2) of the variation in gene diversity ($P = 0.002$), lake size explained 5.0% of the remaining variation ($P = 0.014$) and latitude did not have any significant additional explanatory

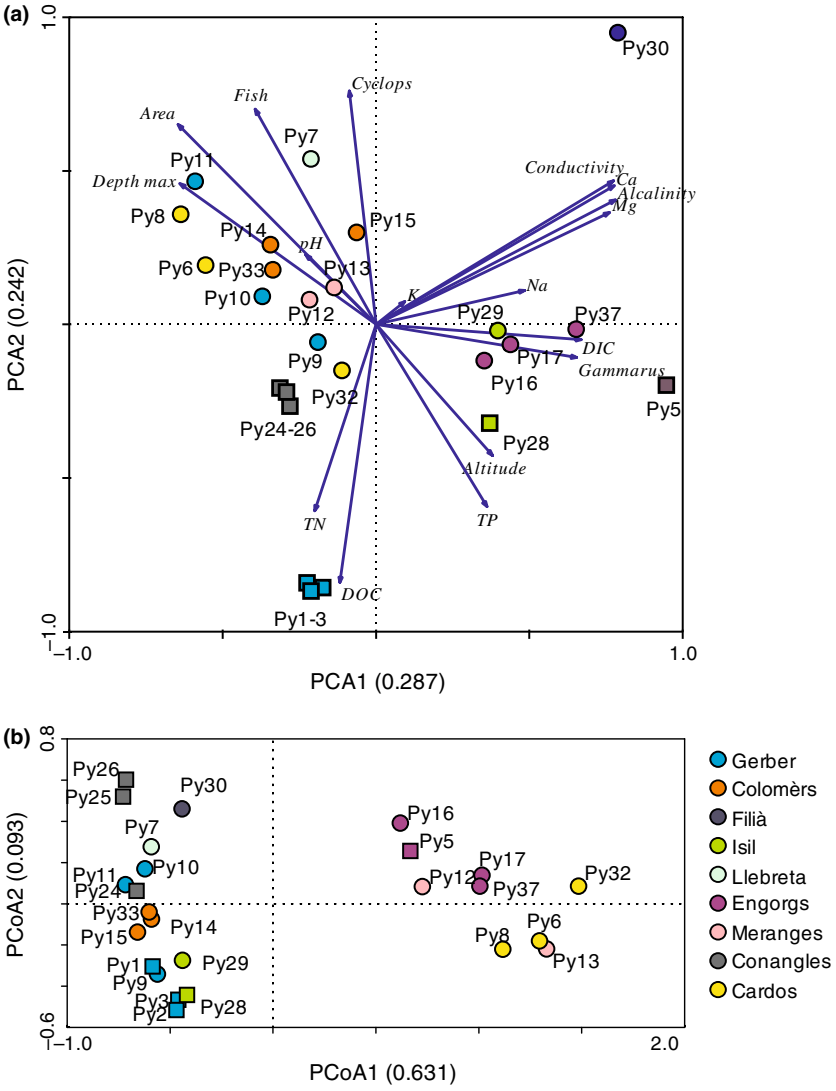


Fig. 4 Genetic similarity of *Daphnia* populations and environmental variation of their habitats. (a) A visual representation of the first two PCA axes summarizing environmental variables and the position of each water body relative to these variables. Ponds are shown as squares, and lakes as circles. (b) Position of each population characterized by the first two axes of a PCo on the genetic structure based on the nine microsatellite loci.

Table 2 Results of the analysis of molecular variance (AMOVA), partitioning the overall genetic variation at nuclear (microsatellites) and mitochondrial DNA (12S) within and among different catchments

	Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	*P*
Microsatellites	Among catchments	6	1511.7	0.84	29.4	<0.001
	Among populations within catchments	23	480.3	0.29	10.2	<0.001
	Within populations	1962	3400.5	1.73	60.4	<0.001
	Total	1991	5392.5	2.87		
12S	Among catchments	6	112.8	0.56	20.5	<0.001
	Among populations within catchments	12	109.0	1.27	46.4	<0.001
	Within populations	108	97.3	0.90	33.0	<0.001
	Total	126	319.2	2.72		

power ($P = 0.552$). When accounting for lake size as a covariable, longitude uniquely explained 53.4 % of the variation in gene diversity ($P = 0.002$). Lake size alone marginally explained 7.8% of gene diversity ($P = 0.048$). A correlation analysis showed that genetic diversity strongly decreased from east to west (Pearson's

Table 3 Results of the redundancy analysis (RDA) to test for isolation by environment and isolation by distance and for their unique contributions (partial RDA) to the overall genetic structure. Each model first shows the overall model performance (global model) and next the sum of the performance of the forward selected variables (FS total) and the contribution of each selected variable to the model when significant. *S*, spatial variables; *E*, environmental variables. *S* + *E* summarizes contribution of all variables, which can be split to the unique contribution of space (*S* | *E*), the unique contribution of environmental variables (*E* | *S*), explained variance that cannot be attributed uniquely to *S* or *E* (shared). The variance not explained by variables included in the model (complement to *S* + *E*) is indicated as unexplained

RDA model	R^2	$R^2_{adj.}$	*P*
S			
Global model	0.748	0.681	0.0010
FS total	0.726	0.671	0.0001
dbMEM1	0.470		0.0001
dbMEM5	0.119		0.0005
dbMEM2	0.080		0.0006
dbMEM4	0.057		0.0005
E			
Global model	0.650	0.30	0.0405
FS total	0.268	0.201	0.0057
Gammarus	0.167		0.0117
Lake area	0.101		0.0430
S + *E*	0.776	0.701	0.0001
S | *E*	0.560	0.470	0.0001
E | *S*	0.060	0	0.0014
Shared	0.156	0.229	
Unexplained	0.224	0.355	

MEM, Moran eigenvector maps.

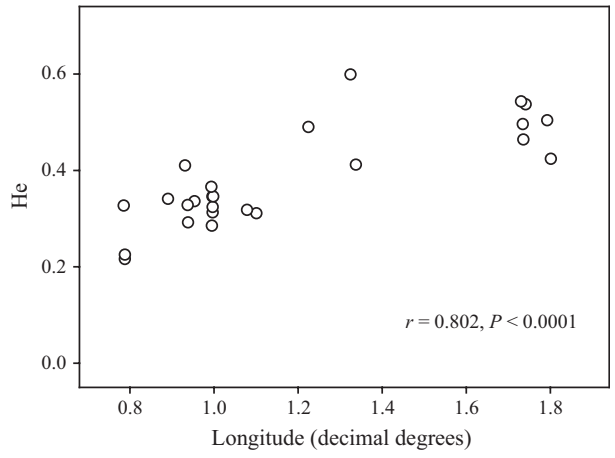


Fig. 5 Plot of eastern Pyrenean *Daphnia longispina* nuclear gene diversity (H_e) vs. lake or pond longitude. Pearson's correlation coefficient is given in the figure.

$r = 0.802$, $P < 0001$), as expected under a model of sequential colonization and persistent founder effects (Fig. 5).

Habitat shifts

Based on the haplotype network (Fig. 2b), we can identify at least two independent habitat shifts from lake to pond in the Pyrenees. They correspond to the two catchments where both lakes and ponds were sampled (Fig. 2b): Py17 (lake) to Py5 (pond), Py9 and Py15 (lakes) to Py1, Py2 and Py3 (ponds). In addition, a dispersal event from pond to lake occurred from Py1-Py3 to Py17. An additional shift also occurred from Py17 (lake) to Zahillo, Sopetón and Taraje in southern Spain (ponds in the Doñana National Park). Further circumstantial evidence for independent habitat shifts from lakes to ponds is apparent from Fig. 4. Ecological differences among spatially close lakes and ponds are typically large (Fig. 4b), whereas genetic distances among the same lake–pond pairs are typically small (Fig. 4b).

Discussion

Genetic diversity

The mitochondrial DNA diversity found in *Daphnia longispina* from 19 water bodies of the Catalan Pyrenees (2700 km^2) is remarkably high, comparable to the diversity found so far in the rest of the studied species' range. The nucleotide diversity in our samples was similar to that found in other parts of Europe and Siberia (Petrusek *et al.* 2008; Giessler & Englbrecht 2009; Thielsch *et al.* 2009; Hamrová *et al.* 2012; Zuykova *et al.* 2013), and the haplotype diversity in our samples (61 haplotypes of 127 sequences from 19 Pyrenean lakes and ponds) was also of the same order of magnitude as the 118 haplotypes from the 417 sequences from the Old World (Fig. 2a). Moreover, nearly all Pyrenean haplotypes likely originated from a single ancestral haplotype and were genetically distinct from most other European haplotypes, except for some lineages apparently derived from the Pyrenean clade. This indicates that the diversification that happened within the Pyrenees is relatively old, likely going back to the early stages of the formation of these lakes (see below). Our results also agree with recent findings of notable differentiation of *D. longispina* in Eastern European mountain lakes (Hamrová *et al.* 2012).

Colonization history and mutation rates

Colonization of the glacial lakes may have occurred directly from external sources to the eastern side of the studied area after the start of the deglaciation of the Pyrenean cirque lakes (ranging between 15 000 and 10 000 years ago; Delmas *et al.* 2008), with subsequent dispersal of *Daphnia* among the newly available glacial

lakes. Alternatively, the colonization might have been from populations with longer history of regional presence in the Pyrenees at altitudes that had remained unglaciated throughout the Late Pleistocene (Gonzalez-Samperiz *et al.* 2006). However, if the diversification had occurred earlier but at lower elevations, with gradual colonization of ponds and lakes higher up the slopes, we would not expect such extreme levels of haplotype endemicity combined with high diversity. It is thus more likely that colonization started only after deglaciation of the upper glacial cirques. Unlike in the Alps, all lowland glacial lakes of the Pyrenees were likely filled with sediment at least ~10 000 years ago (Pallàs *et al.* 2006), which likely favoured the subsequent dispersal of zooplankton mainly through the upper glacial areas.

Establishing a timing of Pyrenean colonization based on the timing of clade expansion requires a reliable estimate of the number of generations per year and of a population mutation rate. The results of the phenology of *D. longispina* coincide closely with those of the same species from the Tatra Mountains (Hamrová *et al.* 2011) and from other observations in the Pyrenees (M. Ventura, personal observation) in that this species has one sexual (resulting in dormant eggs) and one or two asexual generations per year (taking place in the plankton). This species spends the ice-covered period as dormant eggs (encased in ephippia), unlike the other typical *Daphnia* species of the Pyrenees, *Daphnia pulicaria*, which stays in the plankton (Ventura & Catalan 2005). Spending the ice-covered period as dormant eggs implies that each ice-free season the population may be inoculated through hatching of a large number of eggs, resulting in a higher potential for genotype coexistence. On the contrary, in those species or populations overwintering in plankton, it is more likely that clonal selection combined with asexual reproduction results in a lower genetic diversity over the years (Hamrová *et al.* 2011). The observed high levels of genetic diversity at nuclear microsatellite DNA loci match those observed in the same species in other studies (Thielsch *et al.* 2009; Hamrová *et al.* 2011) and are likely a result of overwintering as dormant eggs and the low number of generations per year.

All *Daphnia* in the study lakes and ponds are melanized, supporting the assumption that the UV stress is high at these high-altitude habitats (Laurion *et al.* 2000). Therefore, we expected that the estimated minimal mutation rate would be similar to the one inferred from other UV-stressed *Daphnia* populations (Hebert *et al.* 2002). Given that we have insufficient knowledge on the actual mutation rate of mtDNA in these UV-stressed systems, but that we have knowledge of the number of generations per year and the maximum age of the lakes, we estimated the mutation rates needed to have a colonization of the lakes later than 15 000 years ago (Fig. 3). The figure shows that, for two to three generations per year, there should have been a mutation rate at least higher than 8–12% per million generations. This value is notably higher than previously used for different types of invertebrates including other *Daphnia* (Brower 1994; Costanzo & Taylor 2010), but is close to those experimentally determined for *Daphnia pulex* (Xu *et al.* 2012). Overall, this adds to the evidence that the use of universal molecular clocks and universal mutation rates over wide ranges of organisms or conditions is not warranted (Lynch 2010).

Mechanism explaining the genetic differentiation among populations

Our analyses demonstrate a very strong effect of spatial structure on the genetic data, both at small and at large spatial scales, resulting in a pattern of isolation by distance. After taking environmental covariation into account, the spatial model still explained nearly 50% of the total genetic structure in our nuclear genetic data; the contribution of environmental variation to the genetic data was, on the other hand, negligible (Table 3). The main environmental variable contributing to the model was the presence of *Gammarus*, which was strongly spatially structured: only eastern lakes and ponds host this predator. We thus found no support for isolation by environment and thus for lineage sorting. The very strong spatial structure, even across contrasting environments, indicates that priority effects are long lasting. Given the age of studied habitats, it seems likely that mechanisms further stabilizing initial priority effects played a role, such as local adaptation and build-up of dormant egg bank suggested by the monopolization hypothesis (De Meester *et al.* 2002). Combined with the evidence at the mtDNA of a single founder event for the eastern Pyrenees, a subsequent serial founder effect in a westward direction and an absence of observed gene flow among water bodies (as indicated by the overall paucity of haplotype sharing among lakes and ponds, even within catchments), this strongly suggests that new habitats are typically colonized from the closest neighbouring ones, after which the population may adapt to the local environmental conditions. Therefore, the genetic structure of studied populations strongly suggests that the pattern of isolation by distance is the result of a mechanism of isolation by colonization combined with a serial colonization (Orsini *et al.* 2013a), rather than of isolation-by-dispersal limitation.

Earlier studies have shown the persistent nature of some founder effects (Boileau *et al.* 1992; Orsini *et al.*

2013b). In our study, we can provide an assessment of their scale. We presume the stepwise colonization of catchments within the Pyrenees by *D. longispina* started ~8000 years ago, shortly after deglaciation of the Pyrenees, in a westerly direction, starting from the southernmost parts of the Pyrenees and closest to the climate-buffering influence of the Mediterranean Sea. We thus provide a clear example of priority effects lasting in time (>8000 years old) and occurring at a regional scale (<90 km, the spatial extent of our study) and provide empirical evidence for the strength of the dispersal–geneflow paradox in zooplankton.

Long-distance dispersal

Dispersal is a crucial process for high mountain lake zooplankters, but the previous section indicates that it is most relevant for the colonization phase of water bodies. We found some evidence for dispersal among catchments within the Pyrenees probably representing true gene flow instead of colonization (e.g. a haplotype of Py10 rooted in haplotypes endemic to the Engorgs catchment and a haplotype from the Gerber catchment that dispersed to the Engorgs catchment; Fig. S3, Supporting Information), but these are relatively rare events. In addition, we found evidence of multiple 'ancient' dispersal events in *D. longispina* over thousands of kilometres (which may have occurred in a stepping stone pattern), ending up to almost 5000 km. This contrasts with the much more limited inferred dispersal distances of Ishida and Taylor (2007) on this species (called European *Daphnia rosea* s.l. in their study).

Acknowledgements

Economic support to MV was partially provided by the Spanish Government projects Fundalzoo (CGL2010-14841), Invasivefish (427/2011), the Biodiversity Conservation Plan from ENDESA S. A. (PN 6900014499) and the Grant EST001267 from CSIC to stay at KU Leuven. LDM acknowledges KU Leuven Excellence Center financing PF/2010/07. AP and EH acknowledge the Czech Science Foundation (project no. P506/10/P167) for financial support. Eric Taylor and an anonymous reviewer are acknowledged for their constructive and positive suggestions.

References

Aguilera X, Mergeay J, Wollebrants A, Declerck S, De Meester L (2007) Asexuality and polyploidy in *Daphnia* from the tropical Andes. *Limnology and Oceanography*, **52**, 2079–2088.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. *Ecology*, **89**, 2623–2632.

Boileau MG, Hebert PDN, Schwartz SS (1992) Nonequilibrium gene-frequency divergence – persistent founder effects in natural-populations. *Journal of Evolutionary Biology*, **5**, 25–39.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826–1832.

Borcard D, Legendre P, Gillet F (2011) *Numerical Ecology with R*. Springer, New York, New York, Dordrecht, London, Heidelberg.

Brede N, Thielsch A, Sandrock C *et al.* (2006) Microsatellite markers for European *Daphnia*. *Molecular Ecology Notes*, **6**, 536–539.

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences, USA*, **91**, 6491–6495.

Catalan J, Camarero L, Felip M *et al.* (2006) High mountain lakes: extreme habitats and witnesses of environmental changes. *Limnetica*, **25**, 551–584.

Catalan J, Curtis CJ, Kernan M (2009) Remote European mountain lake ecosystems: regionalisation and ecological status. *Freshwater Biology*, **54**, 2419–2432.

Clegg SM, Degnan SM, Kikkawa J, Moritz C, Estoup A, Owens IPF (2002) Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences, USA*, **99**, 8127–8132.

Costanzo KS, Taylor DJ (2010) Rapid ecological isolation and intermediate genetic divergence in lacustrine cyclic parthenogens. *BMC Evolutionary Biology*, **10**, 166.

De Meester L (1996) Local genetic differentiation and adaptation in freshwater zooplankton populations: patterns and processes. *Ecoscience*, **3**, 385–399.

De Meester L, Gomez A, Okamura B, Schwenk K (2002) The monopolization hypothesis and the dispersal-gene flow paradox in aquatic organisms. *Acta Oecologica-International Journal of Ecology*, **23**, 121–135.

Delmas M, Gunnell Y, Braucher R, Calvet M, Bourles D (2008) Exposure age chronology of the last glaciation in the eastern Pyrenees. *Quaternary Research*, **69**, 231–241.

Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK (2000) High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science*, **289**, 2342–2344.

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.

Duvernell DD, Lindmeier JB, Faust KE, Whitehead A (2008) Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, *Fundulus heteroclitus*. *Molecular Ecology*, **17**, 1344–1360.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Giessler S, Englbrecht CC (2009) Dynamic reticulate evolution in a *Daphnia* multispecies complex. *Journal of Experimental Zoology Part A-Ecological Genetics and Physiology*, **311A**, 530–548.

Gonzalez-Samperiz P, Valero-Garces BL, Moreno A *et al.* (2006) Climate variability in the Spanish Pyrenees during the last 30,000 yr revealed by the El Portalet sequence. *Quaternary Research*, **66**, 38–52.

Hamrová E, Mergeay J, Petrusek A (2011) Strong differences in the clonal variation of two *Daphnia* species from mountain lakes affected by overwintering strategy. *BMC Evolutionary Biology*, **11**, 231.

Hamrová E, Krajicek M, Karanovic T, Černý M, Petrusek A (2012) Congruent patterns of lineage diversity in two species complexes of planktonic crustaceans, *Daphnia longispina* (Cladocera) and *Eucyclops serrulatus* (Copepoda), in East European mountain lakes. *Zoological Journal of the Linnean Society*, **166**, 754–767.

Havel JE, Colbourne JK, Hebert PDN (2000) Reconstructing the history of intercontinental dispersal in *Daphnia lumholtzi* by use of genetic markers. *Limnology and Oceanography*, **45**, 1414–1419.

Hebert PDN, Remigio EA, Colbourne JK, Taylor DJ, Wilson CC (2002) Accelerated molecular evolution in halophilic crustaceans. *Evolution*, **56**, 909–926.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey J (1999) The neutralist, the fly and the selectionist. *Trends in Ecology & Evolution*, **14**, 35–38.

Ishida S, Taylor DJ (2007) Quaternary diversification in a sexual Holarctic zooplankter, *Daphnia galeata*. *Molecular Ecology*, **16**, 569–582.

Laurion I, Ventura M, Catalan J, Psenner R, Summaruga R (2000) Attenuation of ultraviolet radiation in mountain lakes: factors controlling the among- and within-lake variability. *Limnology and Oceanography*, **45**, 1274–1288.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Louette G, De Meester L (2004) Rapid colonization of a newly created habitat by cladocerans and the initial build-up of a *Daphnia*-dominated community. *Hydrobiologia*, **513**, 245–249.

Lynch M (1985) Speciation in the Cladocera. *Verhandlunger der Internationalen Vereinigung für Theoretische und Angewandte Limnologie*, **22**, 3116–3123.

Lynch M (2010) Evolution of the mutation rate. *Trends in Genetics*, **26**, 345–352.

Lynch M, Spitze K (1994) Evolutionary genetics of *Daphnia*. In: *Ecological Genetics* (ed. Real L), pp. 109–128. Princeton University Press, Princeton, New Jersey.

Mergeay J, Verschuren D, Meester L (2006) Invasion of an asexual American water flea clone throughout Africa and rapid displacement of a native sibling species. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2839–2844.

Mergeay J, Aguilera X, Declerck S, Petrusek A, Huyse T, De Meester L (2008) The genetic legacy of polyploid Bolivian *Daphnia*: the tropical Andes as a source for the North and South American *D. pulicaira* complex. *Molecular Ecology*, **17**, 1789–1800.

Nosil P, Egan S, Funk D (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Oksanen J, Blanchet FG, Kindt R *et al.* (2010) vegan: community ecology package. R package version 1.17-3.

Orsini L, Swillen I, Vanoverbeke J, Mergeay J, De Meester L (2013a) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, **22**, 5983–5999.

Orsini L, Mergeay J, Vanoverbeke J, De Meester L (2013b) The role of selection in driving landscape genomic structure of the waterflea *Daphnia magna*. *Molecular Ecology*, **22**, 583–601.

Pallàs R, Rodés A, Braucher R *et al.* (2006) Late Pleistocene and Holocene glaciation in the Pyrenees: a critical review and new evidence from 10Be exposure ages, south-central Pyrenees. *Quaternary Science Reviews*, **25**, 2937–2963.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Peres-Neto PR, Legendre P (2010) Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, **19**, 174–184.

Petrusek A, Černý M, Mergeay J, Schwenk K (2007) *Daphnia* in the Tatra Mountain lakes: multiple colonisation and hidden species diversity revealed by molecular markers. *Fundamental and Applied Limnology*, **169**, 279–291.

Petrusek A, Hobæk A, Nilssen JP *et al.* (2008) A taxonomic reappraisal of the European *Daphnia longispina* complex (Crustacea, Cladocera, Anomopoda). *Zoologica Scripta*, **37**, 507–519.

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences, USA*, **102**, 15942–15947.

Raymond M, Rousset F (1995) GENEPOP (Version-1.2) – population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Salzburger W, Ewing GB, von Haeseler A (2011) The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology*, **20**, 1952–1963.

Schwenk K, Sand A, Boersma M *et al.* (1998) Genetic markers, genealogies and biogeographic patterns in the cladocera. *Aquatic Ecology*, **32**, 37–51.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maxi-

mum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Thielsch A, Brede N, Petrusek A, De Meester L, Schwenk K (2009) Contribution of cyclic parthenogenesis and colonization history to population structure in *Daphnia*. *Molecular Ecology*, **18**, 1616–1628.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Ventura M, Catalan J (2005) Reproduction as one of the main causes of temporal variability in the elemental composition of zooplankton. *Limnology and Oceanography*, **50**, 2043–2056.

Ventura M, Camarero L, Buchaca T, Bartumeus F, Livingstone DM, Catalan J (2000) The main features of seasonal variability in the external forcing and dynamics of a deep mountain lake (Redó, Pyrenees). *Journal of Limnology*, **59**, 97–108.

Wade MJ (2000) Epistasis as a genetic constraint within populations and an accelerant of adaptive divergence among them. In: *Epistasis and the Evolutionary Process* (eds Wolf JB, Brodie EE, Wade MJ), pp. 213–231. Oxford University Press, Oxford, UK.

Wetzel RG (2001) *Limnology: Lake and River Ecosystems*. Academic Press, San Diego, California.

Xu S, Schaack S, Seyfert A, Choi E, Lynch M, Cristescu M (2012) High mutation rates in the mitochondrial genomes of *Daphnia pulex*. *Molecular Biology and Evolution*, **29**, 763–769.

Zuykova EI, Bochkarev NA, Katokhin AV (2013) Molecular genetic identification and phylogeny of *Daphnia* species (Crustacea, Cladocera) from water bodies of the Lake Chany basin. *Russian Journal of Genetics*, **49**, 206–213.

M.V. and J.M. designed the research. M.V., D.B. and A.M. performed the sampling. M.V., E.H. and J.M. performed the genetic analyses. M.V. and J.M. analysed the data. All authors contributed to writing and revising the paper.

Data accessibility

New mtDNA 12S sequences obtained or used in this study are deposited in GenBank, AN: KF977622–KF977696; KJ024375.

Genotypes for microsatellite, as well as a fasta file containing the alignment of the new 12S mtDNA sequences are deposited in the DRYAD databank: doi: 10.5061/dryad.2dh5n

Location and environmental characteristics of studied habitats, descriptive statistics of genetic diversity of microsatellites loci and complementary data analyses and additional results are uploaded as online supplementary material.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Overview of the sampled water bodies, their morphological characteristics, location and connectivity to others in the catchment.

Table S2 List of Accession nos from this study and those from the western Palaearctic retrieved from GenBank.

Table S3 Descriptive statistics of genetic diversity at nDNA, measured at nine microsatellite loci.

Fig. S1 Evidence for a spatially structured demographic expansion.

Fig. S2 Changes in *Daphnia longispina* abundance along the ice-free periods of 2010 and 2011 in Lake Llarg (upper panels) and their body size distribution.

Fig. S3 Haplotype network for variation at the mitochondrial 12S rDNA gene.

Appendix S1 Extended methods on genetic analysis and extended results on genetic diversity and dispersal inferred from mtDNA.

Local and regional founder effects in lake zooplankton persist after thousands of years despite high dispersal potential

M. VENTURA[1,2*], A. PETRUSEK[3], A. MIRÓ[1], E. HAMROVÁ[3], D. BUÑAY[1], L. DE MEESTER[4] AND J. MERGEAY[4,5]

1. Center for Advanced Studies of Blanes, Spanish Research Council (CEAB-CSIC), Biodiversity and Biogeodynamics Group, Blanes 17300, Girona, Catalonia, Spain

2. Institut de Recerca de l'Aigua. Universitat de Barcelona. Av. Diagonal 684, 08034 -Barcelona, Catalonia, Spain

3. Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, CZ-12844 Prague 2, Czech Republic

4. KU Leuven, University of Leuven, Laboratory of Aquatic Ecology, Evolution and Conservation, Deberiotstraat 32, B-3000 Leuven, Belgium

5. Research Institute for Nature and Forest, Gaverstraat 4, B-9500 Geraardsbergen, Belgium

Keywords: Founder effects, *Daphnia*, alpine lakes, Pyrenees, monopolisation hypothesis, dispersal-gene flow paradox

* Corresponding author: Marc Ventura, Center for Advanced Studies of Blanes, Spanish Research Council (CEAB-CSIC), Biodiversity and Biogeodynamics Group, Blanes 17300, Girona, Catalonia, Spain; Fax: +34972337806; e-mail: ventura@ceab.csic.es

Running Title: Persistent founder effects in zooplankton

Type of Article: Original Articles

Supporting information

Table S1. Overview of the sampled water bodies, their morphological characteristics, location and connectivity to others in the catchment. The number of external haplotypes is a proxy for the number of independent colonization events for each water body.

Code	Name	Catchment/region	Sampling year	Lake (L)/ Pond (P)	Fish	Maximum depth m	Lake area ha	Altitude m	Longitude Dec deg.	Latitude Dec deg.	Group	N° external haplotypes	Connectivity
Py1	Bassa Llong Gerber 3	Gerber	2008	P	N	0.6	0.48	2319	0.9966	42.6253	NW	1	<150 m from Py2, Py3
Py2	Bassa Llong Gerber 5	Gerber	2008	P	N	0.1	0.01	2325	0.9959	42.6263	NW	1	Headwater
Py3	Bassa Llong Gerber 6	Gerber	2008	P	N	0.2	0.02	2321	0.9965	42.6261	NW	2	Water from Py2
Py5	Bassa engorgs	Engorgs	2007, 2008	P	N	1	0.07	2547	1.7358	42.4813	SE	1	Water from Py17
Py6	Mariola	Cardós	2008	L	N	46	17.80	2276	1.2243	42.7174	Central	1	c. 8 km from Py8
Py7	Llebreta	St. Nicolau	2008	L	Y	11.5	8.00	1620	0.8903	42.5508	SW	1	c. 6 km from Py33
Py8	Romedo de dalt	Cardós	2008	L	Y	40	11.88	2110	1.3247	42.7060	Central	1	c. 8 km from Py6
Py9	Redo de Gerber	Gerber	2008	L	N	17.9	2.18	2339	0.9980	42.6231	NW	1	Water from Py10
Py10	Illa	Gerber	2007, 2008	L	N	18	2.07	2452	0.9935	42.6184	NW	1	Headwater
Py11	Gerber	Gerber	2008	L	Y	63	14.88	2170	0.9947	42.6307	NW	1	Water from Py9, Py3 and Py1
Py12	Malniu	Meranges	2008	L	Y	13.3	5.46	2250	1.7924	42.4738	SE	2	May have a connection to Py13, 700 m distance, no ridge
Py13	Mal	Meranges	2008	L	Y	3.4	3.69	2260	1.8012	42.4782	SE	1	May have a connection to Py12, 700 m distance, no ridge
Py14	Plan	Colomèrs	2006, 2008	L	Y	13.5	4.95	2188	0.9307	42.6225	NW	1	500 m from Py15, ridge in between
Py15	Manhera	Colomèrs	2006, 2008	L	Y	9.7	1.92	2188	0.9379	42.6243	NW	1	500 m from Py14, ridge in between
Py16	Llarg	Engorgs	2007, 2008	L	N	2	1.81	2490	1.7412	42.4870	SE	1	Water from Py17 and occasionally Py5
Py17	Aparellats de dalt	Engorgs	2007, 2008	L	N	4.4	0.77	2550	1.7342	42.4826	SE	1	Water from Py37
Py24	Bassa baix Redon	Conangles	2005	P	N	2	0.16	2345	0.7851	42.6439	SW	NA	water from Py25
Py25	Bassa mig Redon	Conangles	2005	P	N	0.8	0.07	2380	0.7876	42.6442	SW	NA	20 m from Py26
Py26	Bassa petita Redon	Conangles	2005	P	N	0.4	0.01	2380	0.7880	42.6441	SW	NA	20 m from Py25
Py28	Port d'Aulà	Isil	2006	L	N	3	0.56	2130	1.1005	42.7695	NW	1	Separated from Py29 by a small ridge
Py29	Clavera	Isil	2006	L	N	4	0.48	2230	1.0783	42.7769	NW	NA	Separated from Py28 by a small ridge
Py30	Filia	Filià	2006	L	Y	5.5	1.37	2140	0.9533	42.4512	SW	NA	c. 6 km from Py33
Py32	Becero	Cardós	2006	L	N	5	0.62	2270	1.3374	42.6539	Central	1	c. 5.3 km from Py8
Py33	Clòto de Naut	Colomèrs	2006	L	Y	8	0.91	2330	0.9372	42.5045	NW	1	Water from Py15 and Py14
Py37	Minyons	Engorgs	2008	L	N	1.9	0.93	2580	1.7305	42.4838	SE	NA	Water to Py17

Table S2. List of accession numbers from this study and those from other parts of *D. longispina* range retrieved from GeneBank. Sources abbreviations are T96: (Taylor *et al.* 1996) ; P07: (Petrusek *et al.* 2007) ;P08: (Petrusek et al. 2008) ; T09: (Thielsch *et al.* 2009) ; GE09: (Giessler & Englbrecht 2009) ;H12: (Hamrová *et al.* 2012); Z13: (Zuykova *et al.* 2013).

Name	country	Source	Accession No.
Bassa Llong Gerber 3	Spain	This study	KF977622, KF977623, KF977624
Bassa Llong Gerber 5	Spain	This study	KF977622, KF977623, KF977625
Bassa Llong Gerber 6	Spain	This study	KF977622, KF977623, KF977626, KF977627, KF977628, KF977629
Bassa engorgs	Spain	This study	KF977638, KF977639, KF977640, KF977641, KF977642, KF977643
Mariola	Spain	This study	KF977644, KF977645, KF977646
Llebreta	Spain	This study	KF977647
Romedo de dalt	Spain	This study	KF977644, KF977648
Redo de Gerber	Spain	This study	KF977649, KF977650, KF977651
Illa	Spain	This study	KF977652, KF977653, KF977654, KF977655, KF977656, KF977657, KF977658, KF977659, KF977660
Gerber	Spain	This study	KF977647, KF977661, KF977662, KF977663
Malniu	Spain	This study	KF977664, KF977665, KF977666
Mal	Spain	This study	KF977667, KF977668
Plan	Spain	This study	KF977669, KF977670, KF977671
Manhera	Spain	This study	KF977649, KF977672, KF977673
Llarg d'Engorgs	Spain	This study	KF977633, KF977674, KF977675, KF977676
Aparellats de dalt	Spain	This study	KF977622, KF977630, KF977631, KF977632, KF977633, KF977634, KF977635, KF977636, KF977637, KF977677, KF977678
Port d'Aulà	Spain	This study	KF977649, KF977679, KF977680
Becero	Spain	This study	KF977681
Clòts de Nout	Spain	This study	KF977682
Jezerce "7", Prokletije Mountains	Albania	H12	JX134347
Ligeni i Dashit, Prokletije Mountains	Albania	H12	JX134344
Mondsee	Austria	P08, T09	EF375827, FJ178322
Jugovo, Zelengora Mountains	Bosnia and Herzegovina	H12	JX134328
Veliko, Treskavica Mountains	Bosnia and Herzegovina	H12	JX134325
Ribno Banderishko, Pirin Moutains	Bulgaria	H12	JX134339
Sulzata, Rila	Bulgaria	H12	JX134322
Horní Polka, Bohemian Forest	Czech Republic	P08	EF375837

Jelení, Bohemian Forest	Czech Republic	This study	KF977683
pond near Dobrá, Bohemian Forest	Czech Republic	This study	KF977683
Prášilské Lake, Bohemian Forest	Czech Republic	This study	KF977684
Říjiště, Bohemian Forest	Czech Republic	This study	KF977685
Slapy reservoir	Czech Republic	This study	KF977686
Vranov reservoir	Czech Republic	T09	FJ178341
Žďárské, Bohemian Forest	Czech Republic	P08	EF375835
Brededam	Denmark	P08	EF375836
Erikadam	Denmark	This study	KF977687
Pernillesø	Denmark	P08	EF375837
Store Kobberdam	Denmark	P08	DQ536400
Tana	Ethiopia	P08	EF375828
Bodensee	Germany	P08	EF375829
castle fountain in Heidelberg	Germany	This study	KJ024375
Frankfurt am Main - botanical garden	Germany	P08	EF375839
Zidak pond, Drouzkovice	Czech Republic	P08	EF375834
Hartsee	Germany	GE09	FJ943792
Helgoland	Germany	This study	KF977688
Ismaning	Germany	P08,GE09	EF375838, FJ943787
Klostersee	Germany	GE09	FJ943793
Stechlinsee	Germany	P08	EF375831
Hula	Israel	P08	EF375840
Lago di Campo IV	Italy	T96	U34643
Hridsko, Prokletije Mountains	Montenegro	H12	JX134332
Malo Šiško, Bjelasica mountains,	Montenegro	H12	JX134334
Modro, Durmitor	Montenegro	H12	JX134336
Valovito, Durmitor	Montenegro	H12	JX134327
D'Aova	Morocco	This study	KF977689
Goksjø	Norway	P08,T09	EF375832, FJ178313
Hurdalsvatn	Norway	This study	KF977690
Mildevatn	Norway	P08	EF375841
Molandsvann	Norway	This study	KF977691
Nordfjordvatn	Norway	T09	FJ178330
Storveavatn	Norway	T09	FJ178339
Trollvann	Norway	P08	EF375842
Bucura, Retezat Mountains	Romania	H12	JX134330
Glubokoje	Russia	P08	EF375833

Montenegro, Prokletije ME

Barabinsk pond 1, Chany Lake basin	Siberia	Z13	JN903675
Barabinsk pond 2, Chany Lake basin	Siberia	Z13	JN903680
Inflow of Hargat to Chany Lake	Siberia	Z13	JN903664
Zdvinsk pond, Chany Lake basin	Siberia	Z13	JN903669
Dankovo, Spišská Magura	Slovakia	P07	DQ337938
Jamské, High Tatra Mountains	Slovakia	P07	DQ337932
Malé Čierne Pliesko, High Tatra Mountains	Slovakia	P07	DQ337933
Nižné Jamnické, West Tatra Mountains	Slovakia	P07, T09	DQ337937, FJ178327
Nižné Rakytovské, High Tatra Mountains	Slovakia	P07	DQ337931
Prvé (Dolné) Roháčske, West Tatra Mountains	Slovakia	P07, H12	DQ337935,
Štvrté (Horné) Roháčske, West Tatra Mountains	Slovakia	P07, H12	DQ337936, JX134352
Tretí Roháčske, West Tatra Mountains	Slovakia	P07, H12	DQ337934, DQ337935, JX134353, JX134354,
Vyšné Furkotské, High Tatra Mountains	Slovakia	P07	DQ337929
Vyšné Jamnícke, West Tatra Mountains	Slovakia	H12	JX134349
Vyšné Račkové, West Tatra Mountains	Slovakia	H12, P07	JX134350, JX134351, DQ337934
Vyšné Rakytovské, High Tatra Mountains	Slovakia	P07	DQ337930
Vyšné Satanie, High Tatra Mountains	Slovakia	P07, T09	DQ337939
Laguna del Sopetón, Doñana	Spain	This study	KF977692
pond Dulce, Doñana	Spain	This study	KF977693
pond Taraje, Doñana	Spain	This study	KF977694
Villar del Rey reservoir, Badajoz	Spain	P08,T09	EF375844, FJ178310
Zahillo pond, Doñana	Spain	P08,T09	EF375843, FJ178344
Göteborg, pond in Laerjeholm	Sweden	P08, T09	EF375845, FJ178320
Koarp	Sweden	This study	KF977695, KF977696
Kellersee	Switzerland	P08, T09	EF375827, FJ178322
ponds above Great St. Bernard pass	Switzerland	P08, T09	EF375847, FJ178336, FJ178335
Unterer Arosasee	Switzerland	P08	EF375846

Table S3. Descriptive statistics of genetic diversity at nDNA, measured at 9 microsatellite loci. Abbreviations: N, sample size averaged over loci; P, number of polymorphic loci (out of 9); MLG, clonal richness (number of multilocus genotypes); MLG/N, clonal richness divided by sample size; Div, Simpson's Index of diversity; A, average number of alleles; PA, average number of private alleles; AR_{29}, standardised allelic richness to 29 samples; He, expected heterozygosity; Ho, observed heterozygosity; HWE, Hardy-Weinberg equilibrium; LD, proportion of locus pairs which deviate significantly from linkage equilibrium; F_{is}, average inbreeding coefficient.

Population	**N**	**P**	**MLG**	**MLG/N**	**Div**	**A**	**PA**	**A**	**AR_{29}**	**He**	**Ho**	HWE	LD	**F_{is}**
Py01	32	8	32	1	0.97	2.8	0.00	2.8	2.51	0.313	0.330	0.81	0.06	-0.054
Py02	33	8	33	1	0.97	2.6	0.00	2.6	2.30	0.346	0.350	0.39	0.17	-0.011
Py03	32	8	31	0.97	0.97	2.4	0.00	2.4	2.29	0.324	0.372	0.02	0.09	-0.146
Py05	31	9	31	1	0.97	5.2	0.00	5.2	2.30	0.464	0.434	1.00	0.26	0.065
Py06	32	9	32	1	0.97	4.7	0.22	4.7	4.24	0.490	0.486	0.89	0.14	0.007
Py07	32	6	32	1	0.97	2.6	0.00	2.6	3.77	0.341	0.350	0.20	0.03	-0.027
Py08	31	9	31	1	0.97	6.9	0.78	6.9	2.31	0.599	0.470	1.00	0.69	0.216
Py09	30	8	30	1	0.97	4.2	0.00	4.2	5.59	0.346	0.370	0.55	0.11	-0.071
Py10	31	8	31	1	0.97	5.8	0.22	5.8	3.43	0.366	0.271	1.00	0.63	0.260
Py11	30	5	29	0.97	0.96	3.0	0.00	3.0	4.28	0.285	0.267	1.00	0.03	0.066
Py12	29	8	14	0.48	0.86	3.0	0.11	3.0	2.62	0.504	0.629	0.00	0.37	-0.247
Py13	30	7	30	1	0.97	3.0	0.00	3.0	2.82	0.424	0.463	0.04	0.06	-0.092
Py14	32	9	32	1	0.97	5.6	0.22	5.6	2.81	0.410	0.283	1.00	0.83	0.309
Py15	30	7	30	1	0.97	2.9	0.00	2.9	4.29	0.292	0.263	0.98	0.06	0.098
Py16	32	9	32	1	0.97	5.7	0.00	5.7	2.50	0.537	0.515	0.93	0.17	0.040
Py17	32	9	32	1	0.97	5.6	0.00	5.6	4.55	0.496	0.429	1.00	0.31	0.135
Py24	32	8	32	1	0.97	3.2	0.11	3.2	3.67	0.327	0.321	0.94	0.06	0.019
Py25	60	6	43	0.72	0.97	2.6	0.00	2.6	2.73	0.216	0.211	0.76	0.03	0.025
Py26	40	8	30	0.75	0.95	2.3	0.00	2.3	1.96	0.225	0.212	0.72	0.49	0.058
Py28	32	6	31	0.97	0.97	3.0	0.22	3.0	1.98	0.311	0.302	0.77	0.09	0.028
Py29	31	8	31	1	0.97	4.4	0.22	4.4	2.55	0.318	0.290	1.00	0.26	0.090
Py30	32	7	31	0.97	0.97	2.4	0.00	2.4	3.39	0.336	0.281	1.00	0.17	0.164
Py32	32	8	32	1	0.97	4.0	0.22	4.0	3.17	0.412	0.347	1.00	0.14	0.157
Py33	32	9	32	1	0.97	4.0	0.11	4.0	3.17	0.328	0.340	0.73	0.11	-0.038
Py37	40	9	39	0.98	0.97	5.3	0.11	5.3	4.16	0.543	0.517	0.99	0.29	0.047

Fig. S1. Evidence for a spatially structured demographic expansion. Mismatch distribution plot from mitochondrial data (12S) for populations of the Eastern Pyrenees. Observed frequencies are solid lines, predicted frequencies (spatial expansion model) are dashed lines. The 95% confidence interval (CI) is shown in grey. Tau = 3.561. The sudden expansion model provided a similar fit.

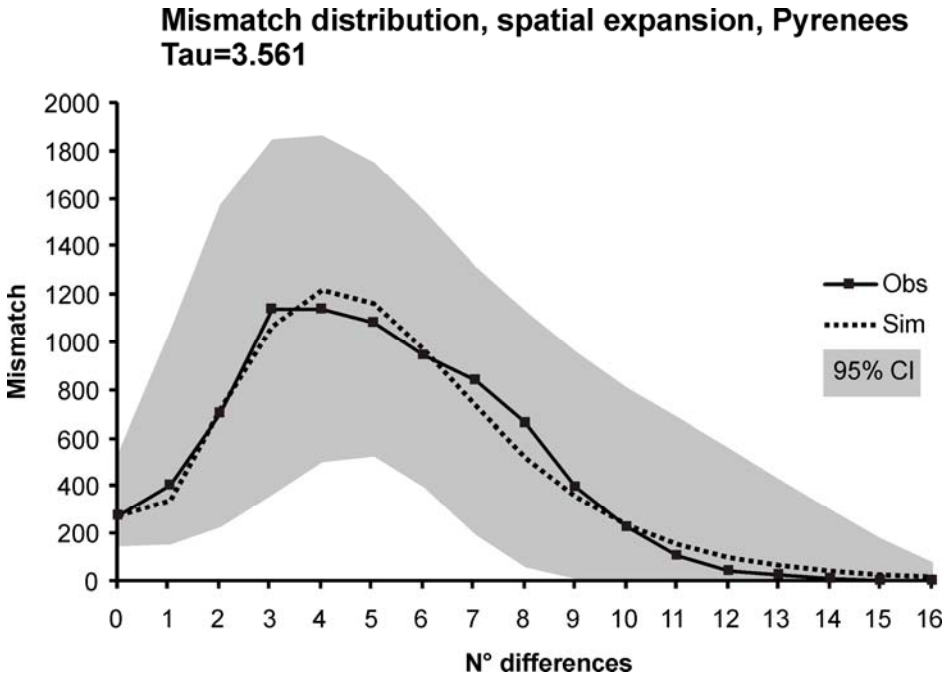


Fig. S2 Changes in *Daphnia longispina* abundance along the ice free periods of 2010 and 2011 in Lake Llarg (upper panels) and their body size distribution. Black circles and bars correspond to reproductive females and white circles and bars to juveniles. Horizontal black bars in the upper panels show the ice-covered period.

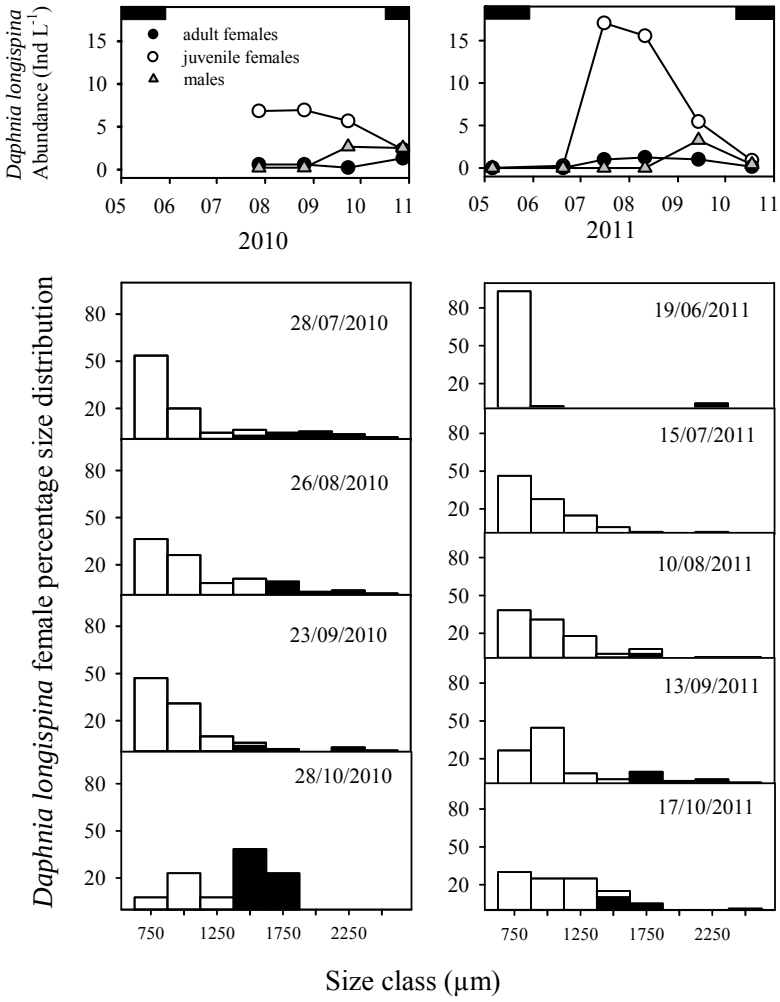
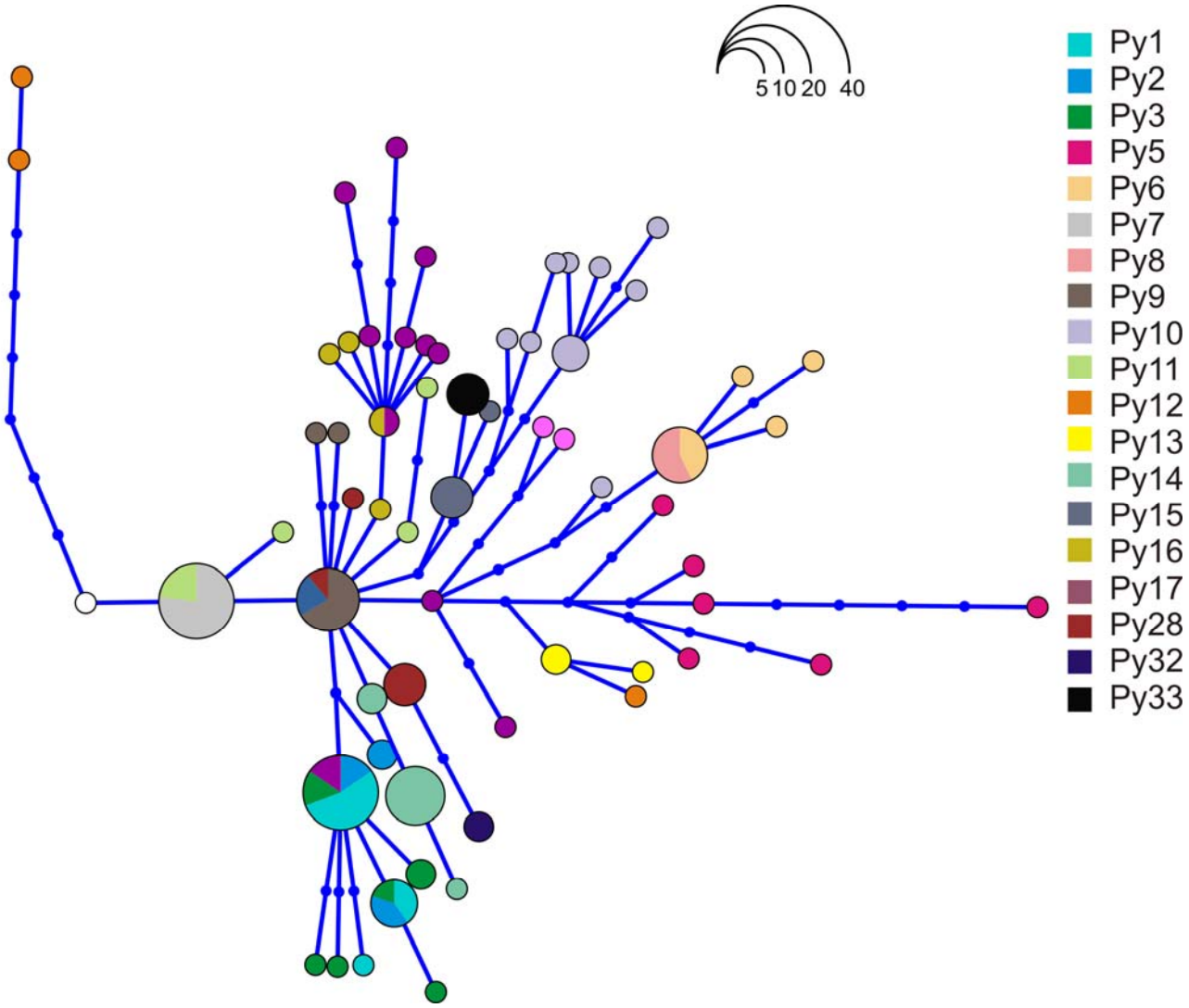


Fig. S3. Haplotype network depicting variation at the mitochondrial gene for 12S rRNA in studied *D. longispina* populations in the Pyrenees. Each circle represents a unique haplotype and its size is proportional to the number of individuals sharing that specific haplotype. Each branch with more than one mutational step is labelled. Studied localities are indicated by a different colour. The locality codes refer to Table S1.



Appendix S1

Extended methods

Extended genetic analysis

We amplified and sequenced *ca*. 540 nucleotides (nt), a long fragment of the mitochondrial 12S rRNA gene (12S) using primers 12S-F (5′-ATGCACTTTCCAGTACATCTAC- 3′) and 12S-R (5′-AAATCGTGCCAGCCGTCGC-3′) (Taylor *et al.* 1996). The reaction mix of the total volume 25 µL contained 1× PCR buffer (Silverstar, Eurogentec), 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.2 µM of each primer, 1 µL of template DNA and 1–2 U *Taq* polymerase. PCR amplifications for 12S involved a denaturing step of 5 min at 95 °C, followed by 30–40 cycles of 45 s at 95 °C, 45 s at 53 °C, 45 s at 72 °C, and a final elongation of 7 min at 72 °C. Purified amplification reactions were sequenced using forward primers on an ABI PRISM 3130 capillary DNA sequencer (Applied Biosystems).

Extended results

Extended genetic diversity

Fourty-five of the 61 haplotypes found in the Pyrenees were singletons; 10 of the remaining 16 sequences occurred in just a single lake. The average number of nucleotide differences between sequences (k) was 4.96, yielding an overall nucleotide diversity (Pi) of 0.0099. The remaining 417 sequences, mostly European but including also samples from Russia, the Middle East and Africa, had 97 variable nucleotides, with 118 haplotypes. One hundred and one of these haplotypes were singletons, 29 of the remaining 52 haplotypes were present in just a single lake and only 4 haplotypes had sequences from lakes belonging to different countries. The average nucleotide difference was 5.334, and nucleotide diversity was 0.01062, indicating that the genetic diversity

within the Pyrenees alone is similar to the so far known genetic diversity across the rest of the species range.

With regards to the molecular variability in nuclear DNA, only three populations (Py3, Py12 and Py13) significantly deviated from Hardy-Weinberg equilibrium with Py3 and Py13 only slightly divergent. Observed clonal (i.e., multi-locus genotype, MLG) richness levels, expressed as MLG/N and Simpson diversity, reached maximal possible values in samples in most populations, i.e., all or all but one multilocus genotypes were different (with MLG/N exceeding 0.9 and usually reaching 1), but marked reductions of clonal richness were observed in Py12, Py25 and Py26. Other measures of genetic diversity (including heterozygosity parameters, standardised allelic richness and inbreeding coefficients) are presented in Table S3. The proportion of private alleles was generally low (0 in 14 of the lakes) and had their maximum in Lake Py08. The proportion of locus pairs deviating significantly from linkage equilibrium was relatively low, with the exception of lakes Py8, Py10 and Py14 that had more than half of the locus pairs with significant values.

Dispersal inferred from mtDNA

Several examples from different catchments may demonstrate these patterns. Within the Gerber catchment, the headwater lake Py10 hosts a diverse but endemic array of related haplotypes, despite discharging directly into Py9, and indirectly into Py11. Similarly, no haplotype sharing was found between the connected Py9 and Py11, whereas both lakes shared haplotypes with lakes or ponds from other catchments (Table S1; Figure S3). Py33 in the Colomèrs catchment receives water from both Py14 and Py15, but we did not find any shared haplotypes with either of them. The sole haplotype found in Py33 is, nevertheless, most related to haplotypes from Py15 (Table S1; Figure S3). Finally, Py5 in the Engorgs catchment is a small pond (c. 600 m^2) that receives water from Py17, 60 m away. Although its haplotypes are more related to those from Py17 and Py16 than to any other water, it is genetically distinct and very diverse, with only endemic haplotypes recorded (Table S1). Py17 also discharges into Py16, and yet it shared only one out of fourteen haplotypes

detected in 20 analysed specimens. Most haplotypes recorded in Py17 have a single ancestor within the Engorgs catchment; however, one haplotype from this lake was shared with the Gerber catchment

References

Giessler S, Englbrecht CC (2009) Dynamic reticulate evolution in a *Daphnia* multispecies complex. *Journal of Experimental Zoology Part A-Ecological Genetics and Physiology,* **311A**, 530-U533.

Hamrová E, Krajicek M, Karanovic T, Černý M, Petrusek A (2012) Congruent patterns of lineage diversity in two species complexes of planktonic crustaceans, *Daphnia longispina* (Cladocera) and *Eucyclops serrulatus* (Copepoda), in East European mountain lakes. *Zoological Journal of the Linnean Society,* **166**, 754-767.

Petrusek A, Černý M, Mergeay J, Schwenk K (2007) Daphnia in the Tatra Mountain lakes: multiple colonisation and hidden species diversity revealed by molecular markers. *Fundamental and Applied Limnology,* **169**, 279-291.

Petrusek A, Hobæk A, Nilssen JP*, et al.* (2008) A taxonomic reappraisal of the European *Daphnia longispina* complex (Crustacea, Cladocera, Anomopoda). *Zoologica Scripta,* **37**, 507-519.

Taylor DJ, Hebert PDN, Colbourne JK (1996) Phylogenetics and evolution of the *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and allozyme variation. *Molecular Phylogenetics and Evolution,* **5**, 495-510.

Thielsch A, Brede N, Petrusek A, De Meester L, Schwenk K (2009) Contribution of cyclic parthenogenesis and colonization history to population structure in *Daphnia*. *Molecular Ecology,* **18**, 1616-1628.

Zuykova EI, Bochkarev NA, Katokhin AV (2013) Molecular genetic identification and phylogeny of Daphnia species (Crustacea, Cladocera) from water bodies of the Lake Chany basin. *Russian Journal of Genetics,* **49**, 206-213.